6-1990

The Comparability of a Field & a Device-Mediated Performance Test for M1 Armor Crewmen

Sylvia Smith
Western Kentucky University

Smith,

Sylvia E.

1990

The Comparability of a Field and a Device-Mediated Performance Test for M1 Armor Crewmen

A Thesis

Presented to

the Department of Psychology

Western Kentucky University

Bowling Green, Kentucky

In Partial Fulfillment

of the Requirements for the Degree

Master of Arts

by

Sylvia E. Smith

June 1990

AUTHORIZATION FOR USE OF THESIS

Permission is hereby

☒ granted to the Western Kentucky University Library to make, or allow to be made photocopies, microfilm or other copies of this thesis for appropriate research or scholarly purposes.

☐ reserved to the author for the making of any copies of this thesis except for brief sections for research or scholarly purposes.

Signed Sylvia E. Smith

Date June 29, 1990

Please place an "X" in the appropriate box.

This form will be filed with the original of the thesis and will control future use of the thesis.

The Comparability of a Field and a

Device-Mediated Performance Test

for M1 Armor Crewmen

Recommended 5/30/90
(Date)

Raymond M. Mendel
Director of Thesis

Elizabeth S. Erffmeyer

Karlene Ball

Approved June 29, 1990
(Date)

Elmer Gray
Dean of the Graduate College

Table of Contents

List of Illustrations

THE COMPARABILITY OF A FIELD AND A DEVICE-MEDIATED PERFORMANCE TEST FOR M1 ARMOR CREWMEN

Sylvia E. Smith April 1990 87 pages

Directed by: R. Mendel, E. Erffmeyer, and K. Ball

Department of Psychology Western Kentucky University

A study was conducted to evaluate the construct validity of four measures of Armor combat performance derived through the Simulation Networking (SIMNET) system. Problems with field testing, such as high cost, low reliability, and lack of realism, has lead the Army to look for alternative methods for soldier evaluation. SIMNET's utility for supplemental training and hardware development has been documented and the device holds promise as a low-cost alternative for soldier evaluation. Performance by 120 M1 tank crews on a SIMNET test was compared to their performance on a field test measuring four critical domains of Armor combat performance: command and control (C^2), communications, position location, and combat driving. Acceptable levels of internal consistency were found for the C^2 and communications dimensions. Some evidence of convergent and discriminant validity were found for these two dimensions through the multitrait-multimethod matrix and analysis of variance procedures. However, more score variance was attributable to undesirable sources (method bias and error) than to desirable sources (convergent and

discriminant validity). Comparing performance on a set of direct-analogue items from the two tests failed to produce greater evidence of convergent and discriminant validity. Soldiers reported performing tasks on SIMNET to be "mostly the same" as performing tasks on the M1 tank. The rank order of soldier's questionnaire responses on the four performance dimensions reflected the rank order of the four correlations of performance on the two tests. The results of this study do not support the construct validity of SIMNET as a performance testing device of critical combat skills. Future research on SIMNET's construct validity should use a SIMNET test and field test with the exact same items and scenarios surrounding the performance of those items. If SIMNET's construct validity still is not supported, better criterion measures should be sought against which SIMNET can be evaluated.

Chapter I

Introduction

Of paramount concern to the U.S. Army in future military conflicts is battle success. The Armor force must be in prime battle-ready condition at all times. Although tanks are complemented by other elements within the combined arms team, such as infantry antitank weapons and the Advanced Attack Helicopter, they are still the most decisive element on the battlefield (Pavitt & Tomich, 1982).

The M1 Abrams is currently the Army's main battle tank. It is an extremely sophisticated weapon system with a high degree of tactical mobility and protected fire power. The M1 tank is a fully tracked, heavily armored, land combat vehicle operated by a four-man crew consisting of the Tank Commander (TC), gunner, driver, and loader. It boasts many high-tech features such as a ballistic computer, laser rangefinder, and gun/turret drive and stabilization system.

In order to maximize the M1 tank's potential, the crew must be highly trained and well suited to their duty positions. Inherent to training and other personnel management concerns is performance evaluation. Other research efforts undertaken by the U.S. Army, for example, examining individual difference determinants of soldier performance, also rely on relevant and accurate measurement of tanker performance.

Although field testing is considered to be the most relevant measure of Armor combat performance, the high cost and other problems associated with field testing have forced the Army to search for other relevant and reliable measures of combat performance. Field testing, as an integral part of field training, still remains the evaluation method of choice. However, the possibility exists to use simulators as relevant, reliable, and low-cost measures of soldier performance for research and personnel decision-making. The Army has already found simulators to be extremely useful for training and equipment development purposes. It stands to reason that their utility could be increased further if they could be used confidently for soldier evaluation.

The purpose of the present research is to examine the construct validity of the Simulation Networking (SIMNET) system for measuring Armor combat performance. Collective crew performance on a SIMNET test and a field test are compared. The field test is a single tank tactical exercise performed on the M1 tank over realistic battle terrain involving combat driving, communications, and enemy tank and infantry engagements. The SIMNET test is a simulator-based platoon tactical exercise which incorporates combat driving, communications, and force-on-force engagements using simulated M1 tank modules and computer generated graphics.

The purpose of this research requires an examination of these two methods of measuring soldier performance. Therefore, field testing as currently used in the Army and associated research is discussed. Also, a description of SIMNET, its utility, especially for soldier evaluation, and pertinent research is reviewed.

It is also necessary to discuss the techniques which will be used for data analysis. The multitrait-multimethod (MTMM) matrix (Campbell & Fiske, 1959) and analysis of variance (ANOVA) technique for use with the matrix (Kavanagh, MacKinney, & Wolins, 1971) will be discussed as these procedures are used to evaluate convergent and discriminant validity in a construct validation paradigm.

Chapter II

Literature Review

Field Testing

The ultimate criterion for measuring the skill of Armor crewmen is performance in combat. However, an actual measure of combat performance is not obtainable. Therefore, the Armor training and testing community supports the use of the field exercise as the most relevant and realistic substitution for actual combat. The vast majority of soldier evaluation takes place within the context of training. Frequently, soldiers are tested to determine if they have reached a desired proficiency level. Three examples follow which illustrate the Army's use of performance evaluation within the context of training.

The Tank Gunnery Tables train and evaluate tank crews on gunnery tasks utilizing actual tanks in the field. However, the Tank Tables fall short of representing a total combat environment with actual opposing forces (OPFOR), not to mention the stress and continued interaction of an actual battle. Closer to this end are the Tank Tactical Tables. They attempt to incorporate the important non-gunnery job behaviors of Armor crewmen within a battle scenario (Field Manual (FM) 17-12-1).

In addition to these field exercises is the Army Training and Evaluation Program (ARTEP). ARTEPs involve field testing from the platoon up to the brigade level. They are usually composed of a number of missions which must be accomplished in a short period of time (3-5 days) and involve OPFOR engagements. ARTEPs are used for training and overall unit evaluation purposes (ARTEP 17-237-10-MTP). However, considering there is little time for repetition of unsatisfactory performance or immediate feedback, they appear more like a test than training (Smith, 1978).

Few examples exist of field testing being performed outside of training. One example is the Soldier Performance Research Project (SPRP) (Graham, Leet, Elliott, Hamill, & Smith, 1989). Phase II of the Armor SPRP tested soldiers on combat critical skills using a single tank tactical exercise on the M1 tank in the field. The purpose of this research was to examine the relationship between mental ability and performance. Supporting this relationship would help justify the additional cost and incentives necessary to recruit soldiers of higher mental ability. Army research utilizing such an elaborate and realistic field test is rare because of the problems generally associated with field testing.

Problems with field testing. Although field testing is perceived as extremely relevant and useful by the Armor testing and training community, evaluating soldier performance through hands-on field exercises places a tremendous demand upon available monetary and manpower resources. It has been estimated that the cost of operating and maintaining a single tank is approximately $535.00 per hour and continues to increase (T. S. Schlecter, personal communication, February 9, 1990). That figure does not include the cost of ammunition or OPFOR vehicles. Personnel requirements are also high for field exercises because of the need for maintenance, OPFOR crews, and safety personnel.

Problems with field testing are further complicated because testing areas are becoming more restricted and less adequate. Land which used to be allocated for field maneuvers is being reduced to allow for additional housing and other developmental projects. The ranges which are still available are less adequate to accommodate the speed and mobility of the M1 tank. The increase in technology from slower, less mobile combat vehicles does not allow for traditional use of ranges as an adequate representation of a modern battlefield (Brown, 1983).

In addition to the decreasing availability and adequacy of resources, at least three other problems hamper the conduct and utility of evaluating soldier performance in the field. First, typical measurement procedures focus on the crew level and are ineffective in providing individual feedback. Therefore, a high performing gunner could be hampered by a poor tank commander (Black & Mitchell, 1986).

Second, the condition of the tanks, range equipment, and weather, can change dramatically over a few days of testing. This makes it difficult to compare performance across tank crews. Also, changes in equipment performance, ammunition characteristics, and firing characteristics make it difficult to compare across engagements within tanks. These problems are serious threats to test reliability (Black & Mitchell, 1986).

Third, typical field exercises do not realistically simulate battlefield conditions. For example, measures of gunnery proficiency are based on tank silhouette targets rather than moving OPFOR vehicles. Although safety and cost warrant the use of these targets, they are not as realistic as the enemy infantry and vehicle targets that are found on the actual battlefield (Black & Mitchell, 1986).

Despite these obstacles to good field testing, it remains the most widely used and accepted means of assessing Armor crewmen performance. If one can obtain the necessary resources, defend the relevancy of the testing, and provide for a meticulous administration of the measure, it is possible to obtain reliable and valid measures of Armor crewmen performance through field testing.

Device-Mediated Testing

The Army's need for highly trained soldiers is increasing due to the technologically advanced equipment being used by soldiers today. However, the problems with field testing make the necessary training and performance evaluation difficult. This dilemma led to the Army's Crawl-Walk-Run training policy which implemented the use of simulators (walk) as a non-resource intensive training medium between classroom instruction (crawl) and field training (run) (Brown, 1983). There has been a great deal of research investigating the validity of using simulators as training devices. However, their ability to produce meaningful scores for performance evaluation has not been adequately investigated. Considering the cost-effectiveness of simulators for training it would stand to reason that their utility could be improved further if they could be used to evaluate soldier performance confidently.

The Army currently utilizes simulators in this capacity. For example, the Unit-Conduct of Fire Trainer (U-COFT) is a high-fidelity whole task trainer which presents computer generated target imagery. It is often used for measuring gunnery proficiency. There has been some empirical research into U-COFT's psychometric properties, specifically reliability. Results have been encouraging. There appears to be an acceptable level of reliability for several of the U-COFT measures (Graham, 1986; DuBois, 1987). However, there has not been any empirical investigations into the validity of U-COFT strictly as a performance testing device at this point.

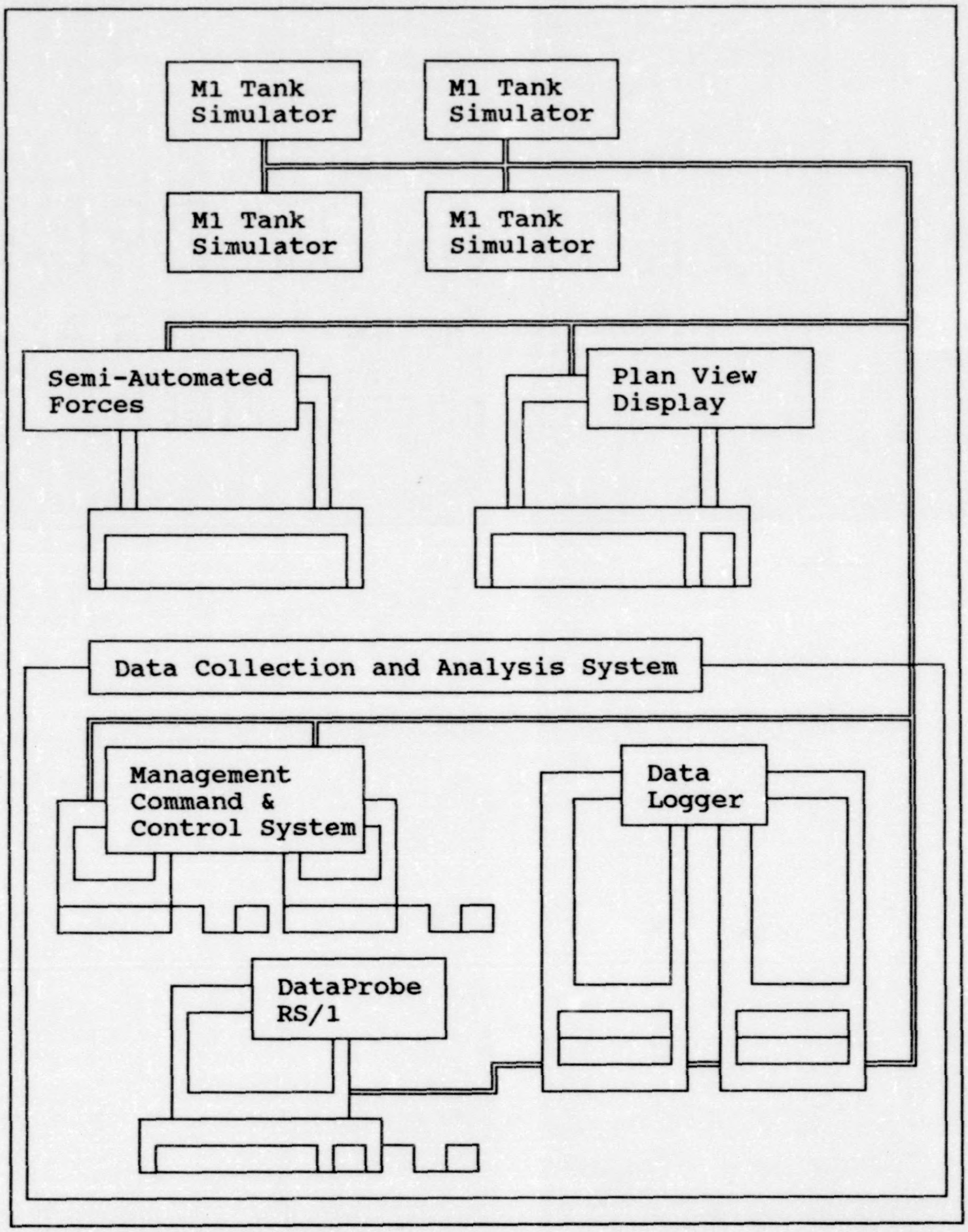
Gunnery performance measures, such as aiming accuracy and speed which are obtainable on U-COFT, are relatively objective when compared to the more tactical tasks of the TC, such as, command and control (C^2) skills and communications. Therefore, even more investigation into reliability and validity are required to have confidence in the scores of tactical training devices such as SIMNET.

The SIMNET System. SIMNET is an advanced research project, currently in its seventh year, sponsored by the Defense Advanced Research Projects Agency (DARPA) in association with the U.S. Army. The goal of the project is to develop the technology to build a large-scale network of interactive combat simulators. This technology is a step above traditional simulators because it goes beyond the single tank crew level to allow fully-manned platoon-, company-, and battalion-level units to fight direct engagements with OPFOR units of the same composition. It has the ability to represent actual military operations through its combined arms environment with the full range of command and control and combat service support elements (Pope, 1987). The outcome of a battle is decided by the strategy and skill of the forces. SIMNET allows soldiers to fight other soldiers, not computers, emphasizing the human interaction aspects of actual battle (Gound & Schwab, 1988).

At each SIMNET installation there is a local area network which connects the computers at that site. Two sites can be connected by a long-haul network so that war-fighting can take place over long distances between battalions. Each simulated exercise requires at least one computer system called a Management Command and Control (MCC) system. The MCC begins and ends the battles under the direction of the Battlemaster. Any number of combat vehicle simulators may be involved. Each represents one combat vehicle, such as a personnel carrier or tank. The MCC system initializes the simulators at the beginning of the exercise with respect to their position and orientation on the terrain, and quantities of ammunition and fuel. Every simulated vehicle periodically reports its position, orientation, and appearance to the other simulators over a local area or long-haul network. The MCC system permits vehicles to be resupplied and repaired, the placement of static gunnery targets, and fire support (Pope, 1987). Figure 1 provides a diagram of the SIMNET network of simulators and data collection and analysis features.

Another feature of SIMNET is a Network Operations and Maintenance (NOM) system which detects and reports hardware failures of the simulators or of the MCC system. The Data Logger records data, such as position location, for research purposes (Pope, 1987). The Plan View Display allows researchers and trainers to have a "bird's eye view" of the battle as it is taking place. The battle can be replayed for review and can be viewed in real time, fast forwarded, or repeated. Gound & Schwab (1988) reported that soldiers and their trainers thought that the Plan View Display was

Figure 1. The SIMNET network.



very helpful in analyzing good and bad points of past performance for training purposes.

Data analysis in SIMNET is supported by a cluster of VAX workstations which extract specific data from a recorded exercise and use those data to produce performance measures. DataProbe[1] and RS/1[2] extract these data and produce descriptive statistics, color graphics, and advanced statistical procedures, such as regression. (Garvey & Radgowski, 1988).

SIMNET development. The development of the SIMNET modules was guided by two philosophies (Chung, Dickens, O'Toole, & Chang, 1988). These philosophies concern the physical fidelity of the modules. Goldstein (1986) defines physical fidelity as the representation of the real world operational equipment. The first developmental philosophy dictated that the behavior of the module should mimic the behavior of the real system as closely as possible. The operational characteristics of the M1 module with respect to present battlefield conditions, such as ammunition loads, vehicle speed, grades being climbed, armor protection, equipment status, fuel capacity, and fuel consumption are continuously updated. The main gun in the M1 module is armed with HEAT and SABOT rounds. It is linked to a

---

[1]DataProbe is a trademark of BBN Software Products Corporation.

[2]RS/1 is a registered trademark of BBN Software Products Corporation.

stabilized laser range finder, gunner's primary sight, and TC's primary sight extension. To add to its realism, sound effects are produced within the tank which simulate engine whine, track movement, weapons fire, and impacting rounds. Also, the crew seats vibrate according to tank speed, road surfaces, and steering and gear changes. The tank is subject to the same type of breakdowns which might occur on an actual battlefield. For example, a tank could slip a track while going up a steep incline (Perceptronics, 1987).

The second developmental philosophy involved the appearance of the simulated tank module and is called selective fidelity. It was either impractical or unnecessary to mirror every aspect of the M1 tank. The four crew stations are represented on the SIMNET M1 module along with many of their controls. The primary restriction on fidelity results from an attempt to limit costs. Therefore, less vital equipment is not reproduced to the physical specifications on the actual tank, yet it still is perceived as realistic. For example, the loader's station has painted knobs in place of some of the controls on the actual tank.

Uses of SIMNET. Currently, there are three major uses of SIMNET. Primarily, it is used to supplement soldier training on combat critical tasks because of the expense and other problems associated with field maneuvers. Secondly, SIMNET is used in the development of new devices and technologies for the M1 tank. SIMNET M1 modules can be easily configured to accommodate a new device or technology allowing an evaluation of the usefulness and appropriateness of the device. This strategy can be particularly cost-effective early in the device-acquisition cycle.

The third use of SIMNET is to evaluate soldier performance. Testing is an integral part of SIMNET with regard to its training function, just as field testing is incorporated into field training. Although tests used for soldier evaluation within the context of training should be well developed and psychometrically sound, there are several reasons why test development for this purpose is generally an easier process than developing tests for research or personnel-decision making purposes.

First, the ability level of the soldiers is known because it is directly related to the level of training they are receiving. Second, the domain of tasks to be included in such a test generally comes directly from training material. Third, results of these tests generally have an impact on training progress, not personnel or policy decisions.

However, the development and psychometric evaluation of performance tests used for research or personnel decision-making purposes is more tedious and critical. The ability level of the subjects must be determined so that tasks will be at the appropriate difficulty level to avoid floor or ceiling effects and allow for variability in performance.

Also, tasks must be carefully selected from a broad domain of possible tasks and must be representative of the skills which are being evaluated. Most importantly, test results can have a far reaching and critical impact, such as demotion or promotion or policy changes. Therefore, the reliability and validity of the test must be investigated and supported before results are used for critical purposes.

The SPRP is an example of research employing a field test and SIMNET test as measures of combat performance. SIMNET was employed because the cost of a field test at the platoon-level was well above the appropriated funds for the research. If SIMNET is to be used for soldier evaluation outside of training, research must support the device's ability to produce test scores which are dependable and meaningful.

So far, research to support the ability of the SIMNET system to evaluate soldier performance reliably and validly has been very limited. This is often the case with simulators employed by the Army, because funds and research efforts have gone primarily to evaluating the training efficiency of these devices. Hoffman and Morrison (1988) examined four computer-based simulation devices that can be used for evaluating M1 gunnery skills. This research examined the usefulness of these devices, including SIMNET, to determine gunnery performance deficiencies. The study also investigated the potential of these devices to test soldier performance and predict performance on the actual equipment.

This study, although limited to gunnery features and subject-matter expert (SME) ratings, offers encouragement for SIMNET's ability to evaluate the C^2 skills of the TC which are related to gunnery tasks, such as giving fire commands. Also, SIMNET's free play format allows for a great variety of scenarios.

Hoffman and Morrison (1988) bring up a very crucial point to consider when correlating simulator performance with performance on real equipment. The correlation is expected to be higher for soldiers who routinely practice on the simulator and on the real equipment than for those who practice on one but not the other. If a soldier is still learning how to operate the simulator equipment and adapt to the device, his performance is more likely to be unstable. Therefore, validity studies which use device and field experienced tankers will maximize the correlation of their performances on the two types of equipment because their test scores will be more stable or reliable.

Limitations of SIMNET. DuBois and Smith (1989) list three ways in which SIMNET is not reflective of a real battle environment. SIMNET M1 modules operate in a closed-hatch mode preventing the TC from viewing the battlefield directly and communicating through hand signals with other units. Also, the TC is only afforded a 64 degree field of vision as opposed to the 360 degree field of vision supplied by the vision blocks on the actual equipment. These restraints result in increased command, control, and communications requirements and reduced navigational and target acquisition capabilities. In an attempt to compensate for some of these difficulties, paper maps are supplied which more accurately simulate the terrain and provide other supplemental navigational aids.

The second major difference between the SIMNET M1 module and the real M1 tank involves the visual cues in the SIMNET computer image generation (CIG) graphics. It is not possible at this point to simulate shadows or many other cues necessary for depth perception. The computer algorithms portray the terrain as if the battle was always taking place at 12:00 noon. Tankers utilize cues from the sun to aid in their orientation and navigation, so these tasks will obviously be more difficult in SIMNET.

The third limitation also involves the CIG graphics. A maximum viewing distance of 3,500 meters is displayed in the module's vision blocks. Obviously, if a tank were on the top of a hill, visibility would be greater than 3,500 meters. Therefore, the horizon soldiers view in SIMNET does not correspond to that of the real world.

Kraemer and Bessemer (1987) found other shortcomings of SIMNET. Differences in the M1 modules fire control system (FCS) did not allow for effective training on gunnery tasks.

Differing ballistics in the SIMNET FCS meant that the soldiers had to aim at a different spot on a target to get a direct hit. They also found that the task of driving was somewhat different in SIMNET. The M1 module did not provide the same type of responsiveness in accelerating, braking, or steering that they were accustomed to in the M1.

It is imperative that SIMNET users take into consideration SIMNET's limitations and design philosophy. SIMNET was designed to train command, control, and communications skills. Other combat critical tasks such as those involved in position location and combat driving have documented differences between their execution in the field and on SIMNET. Only with these limitations in mind can researchers determine if SIMNET is appropriate for their purposes.

Purpose of Research

The main purpose of this research is to gather construct validity evidence for four dimensions of Armor combat performance measured through a test using the SIMNET system. This involves the assessment of reliability and convergent and discriminant validity of the SIMNET performance measures. Collective crew performance on the four dimensions (C^2, communications, position location, and combat driving) is compared to their performance on a field test. Since SIMNET was designed primarily to train C^2 and communications skills it is expected that crew performance on the M1 tank and SIMNET will correlate highly. However, the impact of the differences between the performance of position location and combat driving skills on the two pieces of equipment is less clear.

The Armor community accepts field testing as being the most relevant and realistic measure of combat performance because it measures critical combat duties in the environment and on the equipment that would be used in an actual battle. Therefore, by comparing the psychometric characteristics of scores obtained on the performance dimensions measured by a SIMNET test with those same dimensions measured by a field test, evidence may be obtained to support the use of SIMNET as a reliable and valid measure of tanker combat performance. Research supporting SIMNET's validity and reliability would substantiate SIMNET's use as a surrogate for field testing when high cost or other problems make that method of testing infeasible.

This research uses data collected on a subset of items from the SPRP field and SIMNET tests. The items chosen were those similar across the two measures and comprised the same four major performance dimensions. It is important to note that although items were similar across the two measures, they were not necessarily exactly alike and the circumstances surrounding the performance of those items were different. For example, the field test was a single tank tactical exercise and the SIMNET test was a platoon-level tactical exercise. Also, different types of enemy engagements were used such as the anti-tank guided missile and helicopter which could be simulated in SIMNET, but was impractical for inclusion in the field test. Ideally, tasks and circumstances would be held constant across the two measures in order to maximize the correlation of performance scores (although this would reduce generalizability). However, although some tasks are different they are all still representative of their respective performance dimension as categorized by Army doctrine. Therefore some level of convergence should exist between the two methods.

Assessing SIMNET's construct validity was not the goal of SPRP. However, the rarity of creating such an elaborate and realistic field test for research purposes coupled with the parallel SIMNET test afforded a unique opportunity to examine the issue of the SIMNET test's construct validity.

Construct Validation. The focus of a construct validation approach is on the description of behavior, in this case M1 Armor crew combat performance. Construct validity studies attempt to understand the construct being measured and how well a test or tests measure the construct. Construct validity can not be determined solely through one study, but requires an accumulation of evidence from several different sources (Cascio, 1982). The present research evaluates several sources of evidence for the construct validity of the four dimensions of combat performance: C^2, communications, position location, and combat driving.

One source of evidence inherent in the test development process concerns the content or behavioral domain sampled. Both the field test and the SIMNET test were designed for the SPRP to sample several critical dimensions of tanker performance. These performance dimensions were measured through individual tasks. SMEs reviewed the task lists and as a group categorized the tasks into performance dimensions, which included the four dimensions used in this research and others used in the SPRP. This process helped substantiate the test developers' claim that the tasks were representative of the critical domains of Armor performance.

A second source of evidence for construct validity, one which was assessed through the present research, is internal consistency. Internal consistency is an estimate of the reliability of measurement based on the interrelationship of test items (Ghiselli, Campbell, & Zedeck, 1981). It is used to estimate the degree of homogeneity of the test items within each of the performance dimensions. A high internal consistency indicates the items within a performance dimension are measuring the same overall construct.

A third source of evidence, also assessed through this research, is convergent and discriminant validity. Convergent validation seeks to demonstrate that scores on constructs measured by one method are related to scores on that construct as measured by another method. Discriminant validity is found when scores on a measure of one construct are unrelated to scores on a measure of a different construct when the same and different measurement method is used (Cascio, 1982).

Campbell and Fiske's multitrait-multimethod matrix (1959) displays the correlations between the same trait measured by the same method, different traits measured by the same method, the same trait measured by different methods, and different traits measured by different methods, providing for the assessment of convergent and discriminant validity.

Figure 2 provides an illustration of a MTMM matrix utilizing two methods for measuring four traits. The reliability diagonal (r_1 to r_8) represents internal consistency estimates for the items in each dimension as measured by the same method. The validity diagonal (V_1 to V_4) represents the extent of agreement between two measures of the same trait by different methods. This can provide evidence for convergent validity. The validity coefficients should be significantly different from zero and large enough to encourage further investigation.

Figure 2. An example of a multitrait-multimethod matrix.

		Method 1				Method 2			
	Traits	A_1	B_1	C_1	D_1	A_2	B_2	C_2	D_2
Method 1	A_1	(r_1)							
	B_1	x_{11}	(r_2)						
	C_1	x_{11}	x_{11}	(r_3)					
	D_1	x_{11}	x_{11}	x_{11}	(r_4)				
Method 2	A_2	(V_1)	x_{21}	x_{21}	x_{21}	(r_5)			
	B_2	x_{21}	(V_2)	x_{21}	x_{21}	x_{22}	(r_6)		
	C_2	x_{21}	x_{21}	(V_3)	x_{21}	x_{22}	x_{22}	(r_7)	
	D_2	x_{21}	x_{21}	x_{21}	(V_4)	x_{22}	x_{22}	x_{22}	(r_8)

Note. The parenthesized values (r_1-r_8) in the diagonals of the monomethod triangles are reliability coefficients. The parenthesized values (V_1-V_4) in the heteromethod block are validity coefficients. The solid triangles represent correlations of different traits within the same method. The dashed triangles represent correlations of different traits within different methods.

The obtained validity coefficients expressing the level of convergence between the two measures will be attenuated to the degree that there is unreliability in either measure. Reliability limits validity because the greater the proportion of error variance in the measures, unreliability, the smaller the proportion of true variance available to correlate between the two measures. Although there may be a strong relationship between the constructs measured through SIMNET and the field, high correlations will not be obtained if there is a great deal of error variance present in their measurement (Ghiselli, et al., 1981).

The correction for attenuation formula may be applied to both measures of a construct to estimate the true correlation between the constructs when they are measured without error. This correction estimates the relationship between the measures of the construct if both measures were perfectly reliable. By correcting the convergent validity coefficients of the field and SIMNET performance measures for attenuation, the corrected coefficients would express the true relationship between the constructs as assessed through the two methods. A marked improvement from the uncorrected convergent validity coefficients and the corrected coefficients indicates that increasing the length of the tests is one way to increase the convergent validity coefficients for the dimensions. That is, a SIMNET test may be a good surrogate measure for a field test if there are a sufficient number of items on the test to encourage high reliability.

The next step in evaluating the MTMM matrix is an examination of discriminant validity. Discriminant validity is assessed in three ways. First, the values in the validity diagonal should be higher than the values in the corresponding row or column in which neither trait nor method are in common. Further evidence of discriminant validity is demonstrated if a variable correlates higher with a different method measuring the same trait than with measures of different traits which employ the same method. The third way of examining discriminant validity involves the pattern of trait interrelationships. The same pattern of trait interrelationships should be shown within each of the methods.

Construct validity studies in the literature have found the MTMM matrix to be a very useful technique for gathering evidence of convergent and discriminant validity. For example, Lawler (1967) cited the usefulness of the MTMM matrix for both research and personnel decision-making. His study involved managerial job performance as measured by ratings from three different sources: superior, peer, and self. He found that the supervisor and peer ratings showed good convergent and discriminant validity, but the self-ratings showed little of either. Lawler found the approach allows researchers a much more sophisticated understanding of the criteria because discriminant validity is assessed as well as convergent validity.

Thomson (1970) used the MTMM matrix to examine both criterion and predictor measures of job-related traits of managerial performance. The predictors were ratings by psychologist managers and psychologist supervisors of the subjects' performance at an assessment center. These ratings were designed to be predictive measures of job performance. The criterion measures were supervisor ratings of the subjects' job performance over a period of six to twenty-seven months after they had attended the assessment center. Thomson found the MTMM matrix approach to contribute substantially to the understanding of the sources of invalidity in the measures. The predictor ratings showed high reliabilities and convergent validity indicating that the managers and supervisors agreed on their ratings. There also appeared to be discrimination among traits by the raters. But, the predictive validities were low and showed moderate discriminant validity against the criteria. An examination of the criterion ratings showed low reliabilities and little discrimination on traits by the subjects' supervisors on the job. Therefore, it appeared that either manager or supervisor ratings would make good predictors but the research was compromised due to poor criterion ratings.

Although the MTMM matrix procedure can be very useful for assessing convergent and discriminant validity, it can be extremely awkward when dealing with several traits and/or several methods. In response to this, Kavanagh, MacKinney, and Wolins (1971) proposed an ANOVA technique based on the correlation matrix used in the MTMM approach. This approach transforms the correlation matrix into a more explicit, interpretable, and comparable form. The technique has four advantages. First, it is a more efficient manner to summarize the data. For example, one value is obtained which indicates the amount of convergent validity instead of a number of correlations. Second, there is less judgement involved because the data are explicit and quantifiable. The researcher does not have to try to summarize correlation coefficients of varying degrees into a statement describing all of them. Third, their technique allows for the estimation of method bias (how much the method used influences test scores) and the amount of error variance (unexplainable effects) in the measures. Fourth, the relative strength of the effects can be compared. Thus, the researcher can see how the desirable effects (convergent and discriminant validity) compare with the undesirable effects (method bias and error).

The ANOVA technique enables the estimation of variance components for: a) subjects, which indicates the overall agreement about the subject's performance over methods and traits (convergent validity), b) subject by trait interaction, which indicates the degree of method discrimination on traits by the raters (discriminant validity), c) subject by method interaction, which indicates the amount of method bias, and d) error. Practical information can be gained by comparing the variance contributions from these four sources. It is desirable for the subject and subject by trait variance contributions to be larger than the subject by method and error variance contributions.

Kavanagh et al. (1971) used this ANOVA technique to study managerial performance. They re-evaluated Lawler's

1967 study discussed previously. They found the ANOVA technique to be particularly useful because the variance components could be separated, allowing one to make inferences about the meaning of effects relative to the unexplained variance while controlling for the sample size. By analyzing Lawler's data in this way, Kavanagh et al. initially obtained the same results as did Lawler, good evidence of convergent validity and some evidence of discriminant validity. All four variance components were significant ($p < .001$). The convergent validity variance (subject) component was high, but equally high was the subject by source variance. This indicated a large degree of halo in the ratings. The subject by trait interaction indicated the ordering of subjects differently on different traits, but it was the weakest effect. Also, although all of the F-ratios were significant, they were not very large and the degrees of freedom were very large. The error variance component was the largest of all, suggesting that more variance was due to unknown sources than known sources.

Kavanagh et al. (1971) also used the ANOVA technique to analyze data collected from a longitudinal study of managerial performance described in Kavanagh, MacKinney, and Wolins (1970). The data formed a 60 X 60 matrix consisting of 20 ratings from three different types of managers. The amount of convergent validity can be determined fairly easily by examining the matrix. However, given this large a matrix, the determination of discriminant validity through Campbell and Fiske's (1959) three criteria would be quite difficult. Using the ANOVA model, Kavanagh et al. found each source of variance was significant. Comparing the size of the effects yielded some interesting findings. Although there was good evidence of convergent validity, there was a larger effect due to "halo" (manager X source). Weak discriminant validity (manager X trait) could be explained by the large source bias. They concluded that because there was little discrimination between traits, the number of rating dimensions should be reduced. Additionally, the size of the error variance was approximately equal to the convergent validity and source bias effects indicating that the ratings were subject to unknown sources about as much as they are subject to these known effects.

Direct-Analogue Items

In the present study, the same construct validation process which is applied to the four performance dimensions will also be applied to two of the performance dimensions which represent a smaller subset of items from the field and simulation contexts. This smaller subset of items only contains items which are direct-analogues for one another. That is, although an item may represent the performance domain adequately and was included in the first subset of the SIMNET test, it will only be included in the second subset if the item is the same as an item on the field test.

As was mentioned previously, this type of research generally calls for items measuring a construct to be the same across the measures, thus helping to maximize the convergent validity coefficient obtained between the two measures. There were only direct-analogue items for the dimensions of C^2 and communications. We expect the convergent validity coefficients for the direct-analogue items to be higher than the corresponding coefficients from the full set of items. However, because of a limited number of direct-analogue items, the larger subset of items comprising the four performance dimensions was chosen as the focus of this research.

Soldier's Perceptions

A fourth purpose of this research is to assess soldier's perceptions concerning the similarity of task performance on SIMNET compared to the M1 tank. This addresses the face validity of the SIMNET system, that is, the degree to which soldier's perceive performing tasks on SIMNET to be realistic. It is important that soldiers accept the device as being a realistic surrogate for performing their duties on the M1 tank. High ratings on the similarity of task performance on the two devices would help support SIMNET developer's claim that the system has physical fidelity in terms of mimicking the behavior required on the M1 tank for task performance.

It is also interesting to see if the soldier's perceptions of similarity between the two devices corresponds to the correlation of their actual performance on tasks measured through SIMNET and on the M1 in the field. We expect the rank ordering of the convergent validity coefficients on the performance dimensions and the rank ordering of the mean similarity score on the four dimensions will be similar. This would indicate that soldier's perceptions concerning the similarity of task performance on SIMNET compared with the M1 tank corresponds to their actual performance. It is important for soldiers to believe SIMNET is realistic or else they will not see the value in training or being tested on the device and will lack motivation to perform to their capacity. One explanation for low convergence between performance on the two devices could be that soldiers do not feel that the performance of tasks on SIMNET is comparable to the M1 tank for certain tasks.

In summary, this research attempts to:

1. Gather construct validity evidence for four SIMNET based performance dimensions of the Armor task domain by assessing their internal consistency and convergent and discriminant validity.

2. Determine the extent to which a set of direct-analogue items for two performance dimensions shows a greater construct validity than a set of items with apparently less fidelity.

3. Assess soldier's perceptions concerning the

similarity of performing tasks on SIMNET compared to the M1 tank and how the perceptions correspond to the magnitude of the correlation of task performance on the two devices.

Chapter III

Method

Research Participants

One-hundred twenty TCs and 120 drivers, MOS 19k (M1 Armor crewmen) soldiers, were selected from five Continental U.S. (CONUS) divisions. Of the TCs, 46 were of the rank Sergeant (SGT), 71 Staff Sergeant (SSG), and 3 Sergeant First Class (SFC). Of the drivers, 110 were Specialist (SPC), 5 SGT, and 5 Private First Class (PFC).

Soldier Selection

The criteria for soldier selection were driven by the main purpose of the SPRP and unrelated to the present research, that is, determining the relationship between mental ability and soldier performance. TCs and drivers were selected on the basis of their classification into one of four levels of mental category as defined by the Armed Forces Qualification Test (AFQT). Soldiers were selected from their unit rosters with an equal number of TCs and drivers from each of the mental categories from each division. Soldier training level was also held constant by maintaining the same ratio of Staff Sergeants (SSG) to Sergeants (SGT) in each cell. The 16-cell (4 levels of TC mental category x 4 levels of driver mental category) design was filled systematically by TC and driver pairs distributed equally from their units to counterbalance unit training effects. Testing took place by division, so TCs and drivers were paired from the same division, but not the same unit.

Research Confederates

The TC and driver pairs to be evaluated were combined with confederate gunners and loaders to form reconstituted tank crews for both the field and SIMNET tests. A total of 20 confederates participated in the testing. Twelve occupied the gunner and loader positions for the field test tank crews. Eight confederates participated in the SIMNET testing. Two of the confederates occupied the gunner and loader positions in the test crew module and the other six were TCs and drivers for the three other tank simulators in the four tank platoon. All of the confederates were either of the rank SGT or SSG, and were currently either instructors or serving on Armor units at Fort Knox, KY. The confederates participated in a two-week training program in which they were trained to act in an appropriate and standardized manner for all test crews. In an effort to promote consistency, visual cues were used to prompt the confederates to make pre-determined responses, such as the sighting of an OPFOR vehicle. In addition to the classroom training for the confederates, many practice repetitions were completed on the actual test course. Approximately 100 additional support personnel were required to construct and execute the field and SIMNET tests.

Instruments

Field Test. The field test was designed to measure M1 tank crew combat performance through a single tank tactical exercise performed in a realistic combat field setting. The test was developed in three separate stages. First, the mission was defined in which the test crew was to prepare their tank for combat and traverse a 12 kilometer road encircling a territory which has been recaptured from the enemy. They were to secure the area by searching for and destroying any remaining enemy stragglers within the territory.

Second, eight events were selected from Field Manual 17-12-1, the basic field training guide for soldiers at this level, which were likely to take place within the prescribed mission. These events were distributed around the test course and referred to as stations. Station 1 required the crew to prepare their tank for combat. Stations 2-8 involved encounters with friendly military police and engagements with enemy infantry and Armor forces (Appendix A).

In the third phase of the field test development, tasks were selected, also from Field Manual 17-12-1. The tasks were selected on the basis of their compatibility with the events, ease of observation, and ease of scoring. Although this test measures performance at the crew level, the tasks emphasized the skills of the TC and driver in particular.

The gunner and loader positions were staffed with test confederates trained to act in a standardized. Thus, any performance variance across crews is attributable to the TC and gunner.

Eighty-seven tasks were selected from the field test for use in this research. Tasks were selected which were representative of the four major dimensions of performance: C^2, communications, position location, and combat driving. The tasks had been previously classified by SME's into the major categories of Armor combat performance through the use of Field Manual 17-12-1 (Appendix A).

A second subset of forty-three items was chosen for analysis from the above 87 tasks. These were tasks for which their was a direct-analogue in the SIMNET test. That is, the item had to be nearly identical in both the field and SIMNET scenarios (Appendix A). The selection of these tasks was done by the author and supported by Army Research Institute personnel.

Tasks were scored dichotomously (Pass/Fail). Composite scores for each dimension were determined by the sum of "passing tasks" and expressed as a percentage.

To best approximate actual combat, the realism and stress created by the test were of prime concern. Of course, safety concerns required the use of blank ammunition, so Hoffman charges were used to simulate tank fire. OPFOR vehicles used simulated smoke from fire extinguishers to indicate a hit and destruction of their vehicle. It was pre-determined how many rounds had to be fired at each target before a hit was obtained and this was held constant for each crew tested. At one point during the test, the confederate loader acted as though he had been wounded by gunfire and squirted fake blood about the inside of the tank and on the TC. This was done in an attempt to create the stress that is present in an actual combat situation. The typical reaction by the TCs was panic until they realized it was part of the test and not an actual casualty. They then attended to the "dead" confederate loader in the prescribed manner and reconfigured for a three-man crew.

An effort was made throughout the design and implementation of the field test to eliminate or control as many of the problems associated with field testing as possible. The test was more realistic than most field tests because of the OPFOR VISMODS, i.e., Sheridan tanks visually modified to resemble Soviet tanks, as opposed to using tank silhouette targets. Fire extinguishers and Hoffman charges also added to the realism.

SIMNET Test. The SIMNET test was also designed to measure M1 tank crew combat performance through a platoon-level tank tactical exercise. The test required the test crew (with gunner and loader confederates) to occupy one SIMNET M1 module while three other confederate crews (TC and driver positions only) occupied the three other M1 modules in the platoon. The test was developed by the same three phase method as the field test. The test crew's mission was to join a platoon as a wingman and scout a combat zone. The eight events consisted of enemy engagements from tanks and helicopters and combat driving platoon formations.

One hundred and twenty-eight tasks were selected from the SIMNET test for inclusion in this research. The tasks selected were those which comprised the four major performance dimensions of interest. These tasks had previously been categorized by SME's into the performance dimensions. A second subset of forty-three direct-analogue items was also selected (Appendix B).

Task Similarity Questionnaire (TSQ). A ten item questionnaire was developed to assess the soldier's perceptions of the similarity of performing tasks on SIMNET compared with the actual M1 tank. The purpose of this questionnaire was to assess the face validity of the SIMNET system. That is, does SIMNET appear to the soldier's to require the same processes to perform tasks as the M1 tank. That is, do the soldiers feel as though they are in an M1 tank carrying out their duty functions. Responses were provided on a five-point Likert scale ranging from (1) completely different to (5) completely same performing the task (Figure 3).

Questionnaire items were composited to form the four performance dimensions. Items 1 and 2 comprise communications, items three and four comprise C^2, items five and six comprise position location and combat driving is represented by item seven alone. Items 8, 9, and 10 were not used in the composites because they represented tasks which comprise other dimensions of performance that were used in the Soldier Performance Research Project.

Figure 3. Task Similarity Questionnaire.

Task Similarity Questionnaire

This questionnaire asks you to compare the performance of the following activities on SIMNET with the performance of those activities on an actual M1 tank in the field. Please indicate your response by circling the appropriate number by each task.

Task	Task Similarity in SIMNET Compared to M1 Tank				
	Performed Differently.............Performed Same				
	Completely Different	Mostly	Neutral	Mostly	Completely Same
	1	2	3	4	5
1. Giving combat reports	1	2	3	4	5
2. Following radio procedures	1	2	3	4	5
3. Commanding the crew	1	2	3	4	5
4. Directing engagements	1	2	3	4	5
5. Determining position location	1	2	3	4	5
6. Map reading	1	2	3	4	5
7. Combat driving	1	2	3	4	5
8. Security	1	2	3	4	5
9. Call and adjust indirect fire	1	2	3	4	5
10. Troop leading procedures	1	2	3	4	5

Besides examining the soldiers' opinions on task similarity, it is also desirable to see if their perceptions correspond to the correlation of their actual field and SIMNET performance on the four dimensions. Therefore, the rank order of the TSQ performance dimension composites will be compared to the rank order of the convergent validity coefficients for the performance dimensions.

Equipment

The field testing was conducted on Wilcox Range, Fort Knox, KY. Six M1 tanks were used so that more than one crew could be on the test course at a time. The SIMNET test required four M1 modules. The Plan View Display, DataLogger, and a transmitter to hear conversation within the test module were used for data collection purposes.

Data Collection

The field test data were gathered on task checklists and compiled onto a score sheet. Most of the field test data collectors were stationed at an observation area overlooking the field test site. They monitored the radios that the test tank used to send reports and monitored the intercoms which relayed all of the conversations from within the test tank. Recordings were made so that the data collected could be checked later, if needed. Additional data were gathered by the confederate loaders and drivers within the test tanks and by observers along the test course.

SIMNET data were also gathered by multiple data collectors using checklists and then compiled onto a master score sheet. The majority of data collection was accomplished by monitoring the tank intercom and the test crew's radio transmissions. Also monitored were the Plan View Display giving a "bird's eye view" of the test crew's progression. SIMNET data were also collected by viewing the shadowbox which provided the same line of sight as the driver and TC within the test module. The DataLogger was used to corroborate some of the data.

Pilot Test

Field and SIMNET pilot tests were conducted to insure the standardization of test administration for all of the test crews. The consistency of the confederates and the adequacy of the data collection methods were of prime concern. For the pilot testing, the field test confederates were the test crews for the SIMNET test and the SIMNET test confederates were the test crews for the field test. The results of the pilot test indicated 1) the number of crews that could feasibly be tested in a day, 2) that the confederates behavior was standardized across test crews and, 3) that data could be collected adequately through the proposed methods.

Procedure

The soldiers arrived at the test site one division at a time the night before field testing. They set up camp in groups of 16, were briefed on the research, and given a copy of the unit Standard Operating Procedure to study to help refresh their training. The following morning the soldiers were introduced, one pair at a time, to the confederate gunner and loader who were designated as members of their crew. Approximately nine crews were tested each day. The complete testing for a division took place over four days with one-third of the research participants being tested in the field on the first day. The second day, soldiers tested in the field previously were tested on SIMNET and the next third of the participants were tested in the field. This pattern was replicated on day three and on day four the remaining third finished their testing on SIMNET.

Before SIMNET testing, the research participants received training to familiarize them with the SIMNET system. Training consisted of about 20 minutes of classroom instruction, 40 minutes of hands-on experience with a familiarization course where the crews maneuvered cross country, engaged targets, and learned the consequences of running into objects, and 60 minutes of formal training in which the instructor gave limited assistance. At the end of training, the soldiers were tested on a 30 minute certification course in which they had to complete all of the critical tasks satisfactorily or else be retested until they did. The tasks included on the certification course were more difficult than those included in the actual SIMNET test. The purpose of this was to insure that the soldiers were familiar enough with the SIMNET system to perform the tasks that would be on the test.

The soldiers were tested a crew at a time on SIMNET. First, they were briefed on their mission and introduced to the confederate gunner and loader. Testing lasted about an hour and the participants were debriefed at the end. At this time, they were asked to fill out the Task Similarity Questionnaire and some other measures used in the SPRP.

Chapter IV

Results

Field and SIMNET Measures

Descriptive Statistics. The means and standard deviations for the performance dimensions of the field and SIMNET tests are presented in Table 1. All of the performance dimensions appear to contain sufficient variance, equal to or above 10%, to discriminate among the test crews.

Table 1.

Means and standard deviations for the performance dimension composites of the field and SIMNET tests.

PERFORMANCE DIMENSION		FIELD % Correct	SIMNET % Correct
Command and Control	Mean	.55	.67
	SD	.14	.15
Communications	Mean	.53	.49
	SD	.13	.10
Position Location	Mean	.40	.51
	SD	.14	.12
Combat Driving	Mean	.80	.74
	SD	.18	.14
Average	Mean	.57	.60
	SD	.15	.13

Note. N = 120. SD = standard deviation.

The mean level of performance is similar across the field and SIMNET tests with crews averaging 57% of the total test items correct on the field test and 60% for the SIMNET test. Levels of performance appear to be fairly comparable across the dimensions and methods except for the combat driving dimension for which performance was considerably higher on both methods.

Internal Consistency. The results of the internal consistency reliability calculations for the performance dimensions are presented in Table 2.

The Kudor-Richardson Formula 20 (KR-20) for dichotomously scored variables was used to calculate internal consistency (Ghiselli et al., 1981). Acceptable levels of internal consistency were found for the dimensions of C^2 and communications for both the field and SIMNET tests. However, the position location and combat driving dimensions show poor internal consistency. The levels of internal consistency comparing the dimensions across methods are quite similar, except for combat driving which obtained a moderate level of .65 on the SIMNET test but was the lowest of all the dimensions at .26 for the field test.

The Spearman Brown correction formula was used to estimate the level of internal consistency that would be obtained if each dimension had 56 items. A marked improvement in the reliability levels is evident for the position location and combat driving dimensions for both the field and SIMNET tests. The corrected reliabilities are much more comparable across dimensions and methods than the uncorrected reliabilities. This indicates that the position location and combat driving dimensions probably do not have inherently lower internal consistency than the C^2 and communications dimensions.

Table 2.

Internal consistency reliability, corrected reliability, number of scale items, and suggested number of items to obtain a KR-20 = .80 for the performance dimension composites of the field and SIMNET tests.

PERFORMANCE DIMENSION		FIELD	SIMNET
Command and Control	KR-20	.71	.82
	$r_{cc'}$	.82	.89
	N of Items	31	32
	Suggested N	51	28
Communications	KR-20	.78	.76
	$r_{cc'}$	.84	.76
	N of Items	38	56
	Suggested N	43	71
Position Location	KR-20	.39	.41
	$r_{cc'}$	.78	.68
	N of Items	10	18
	Suggested N	63	104
Combat Driving	KR-20	.26	.65
	$r_{cc'}$	.77	.83
	N of Items	6	22
	Suggested N	68	47
Mean	$\overline{\text{KR-20}}$	.54	.66
	$\overline{r_{cc'}}$	.80	.79
	N of Items	21	32
	Suggested N	56	62

Note. N = 120. $r_{cc'}$ = Spearman Brown corrected reliability for a 56 item measure.

The results found in this research appear to be due to the fact that fewer items were used for these dimensions in the field and SIMNET tests. This suggests we can construct tests with a larger number of items assessing position location and combat driving skills on SIMNET that would have acceptable levels of internal consistency.

The Spearman Brown correction formula was again used to estimate the number of items that is necessary for each dimension for both a field and SIMNET measure in order to obtain a KR-20 of .80 (Table 2). Substantially fewer numbers of items are needed to test C^2 skills than the other three dimensions on SIMNET. A large number of items is required to reliably measure position location skills. It is possible that the position location tasks used in this SIMNET test are multi-dimensional.

Intercorrelations Among Dimensions. Next, the intercorrelations between the dimensions are presented in the monomethod triangles of the MTMM matrix (Table 3). The correlation between C^2 and communications is the highest on the field test ($p < .01$), yet it is low on the SIMNET test. The correlation between position location and combat driving is the highest on the SIMNET test, yet low on the field test. Clearly the intercorrelations among the dimensions show a different pattern on the two testing methods. This indicates that the factor structures for these dimensions are different when tested on SIMNET than in the field.

Multitrait-Multimethod Matrix. Significant convergent validity coefficients are shown for the dimensions of C^2 and communications ($p < .05$ and $p < .01$, respectively) (Table 3). Corrected for unreliability, the correlations for C^2 and communications are .26 and .56, respectively.

Although significant, the convergent validity coefficients for C^2 and communications are low and only slightly improved when corrected for unreliability in both measures. This indicates a low degree of convergence between the corresponding measures tapped by these two methods.

The convergent validity coefficients for position location and combat driving are not significant indicating no convergence by the two methods on these dimensions. The low levels of reliability for these dimensions are a likely contributor to the insignificant convergent validity correlations for these dimensions.

At a minimum, the convergent validity coefficients should be significant in order to encourage further investigation. Therefore, the assessment of discriminant validity will only be performed for C^2 and communications. Discriminant validity is assessed in three ways. First, the value in the validity diagonal should be higher than the values in its row or column in which neither dimension nor method are in common. This involves the correlations within the heterotrait-heteromethod block. Based on the criterion, discriminant validity is evident for all six comparisons for both C^2 and communications.

Table 3.

Multitrait-multimethod correlation matrix for the field and SIMNET tests.

Method/ trait	FIELD C^2	FIELD CO	FIELD PL	FIELD CD	SIMNET C^2	SIMNET CO	SIMNET PL	SIMNET CD
FIELD								
C^2								
CO	.40**							
PL	.24**	.21*						
CD	.08	.02	.11					
SIMNET								
C^2	**.20***	$\underline{.03}$	.10	.11				
CO	.03	**.43****	.15*	$\underline{.01}$	$\underline{.08}$			
PL	.12	.10	**.09**	$\underline{.07}$	.07	.41**		
CD	.13	.03	.04	$\underline{\mathbf{.03}}$	.13	.16*	.58**	

Note. The correlations in the heteromethod block with boldface type are convergent validity coefficients. The solid triangles are heterotrait-monomethod correlations. The dashed triangles are heterotrait-heteromethod correlations. CC = Command and Control, CO = Communications, PL = Position Location, CD = Combat Driving.
* $p < .05$, ** $p < .01$.
Underlined correlations are negative.

Next, the convergent validity coefficients should be higher than the correlations between different dimensions within the same method. This involves the heterotrait-monomethod triangles. The convergent validity coefficient for C^2 is higher in two of the three comparisons. Specifically, C^2 correlates higher with communications and position location on the field test. The convergent validity coefficient for communications is higher for all three comparisons.

Next, discriminant validity is shown if the pattern of dimension interrelationships is the same for each method. As stated earlier, Table 3 reveals a very different pattern of dimension interrelationships for the field test than for the SIMNET test.

Analysis of Variance. The ANOVA technique was used to summarize the information presented in the multitrait-multimethod matrix and present it in a more precise form. The ANOVA results for the four sources of variance are presented in Table 4. The results of the significance tests on the main effect and interactions indicate that each source of variance is significant ($p < .001$).

Table 4.

Analysis of variance for field and SIMNET tests.

Source	df	MS	F	Variance Component
Subject (S) (test crews)	119	1.91	2.79*	.1850
S X Dimension	357	.94	1.39*	.1131
S X Method	119	1.29	1.88*	.1351
Error	357	.68		.6923

Note. * $p < .001$.

The significant Subject effect indicates that there is differentiation among test crews attributable to the test (convergent validity). There is also differential ordering of test crews on the different performance dimensions as indicated by the significant Subject X Dimension interaction (discriminant validity), however this is less than the size of the Subject effect. The Subject X Method interaction is also significant and is greater than that for the Subject X Dimension interaction indicating more method variance than discriminant validity in the scores. Therefore, it appears there is some evidence of convergent validity, slight evidence of discriminant validity, and a large degree of method bias. An examination of the variance components (Table 4) shows a large error variance, substantially larger than any other effect. This indicates that more variance is attributable to unknown sources than to desirable sources.

Direct-Analogue Items

Descriptive Statistics. The means and standard deviations for the performance dimensions of the direct-analogue items of the field and SIMNET tests are presented in Table 5. Both performance dimensions using both methods appear to contain sufficient variance to discriminate among the test crews. The average level of performance is similar across tests with crews averaging 58% of the total test items correct on the field test and 60% for the SIMNET test.

Levels of performance are similar across dimensions and methods.

Compared to the larger subset of items, scores on the performance dimensions comprised by the direct-analogue items are very similar. There is somewhat greater variability in performance for the direct-analogue items as evidenced by the larger standard deviations for both dimensions by both methods.

Table 5.

Means and standard deviations for the performance dimension composites of the direct-analogue items from the field and SIMNET tests.

PERFORMANCE DIMENSION		FIELD % Correct	SIMNET % Correct
Command and Control	Mean	.62	.63
	SD	.18	.19
Communications	Mean	.55	.57
	SD	.15	.13
Average	Mean	.58	.60
	SD	.17	.16

Note. N = 120. SD = standard deviation.

Internal Consistency. The results of the internal consistency reliability calculations for the performance dimensions are presented in Table 6. All of the uncorrected reliability coefficients are lower than those obtained with the larger subset of items. This is due to the fewer number of direct-analogue items comprising the dimensions compared to the first subset of items. When the correction formula is applied, all of the corrected reliability coefficients are higher with the direct-analogue items than the larger set of items. Also, compared to the larger subset of items, a fewer number of items are required to achieve a test reliability of .80.

Table 6.

Internal consistency reliability, corrected reliability, number of scale items, and suggested number of items for KR-20 = .80 for the performance dimension composites comprised of the direct-analogue items of the field and SIMNET tests.

PERFORMANCE DIMENSION		FIELD	SIMNET
Command and Control	KR-20	.68	.79
	$r_{cc'}$	.87	.92
	N of Items	18	18
	Suggested N	34	19
Communications	KR-20	.72	.64
	$r_{cc'}$	.85	.80
	N of Items	25	25
	Suggested N	39	56
Mean	$\overline{\text{KR-20}}$	.70	.71
	$\overline{r_{cc'}}$	.86	.86
	N of Items	22	22
	Suggested N	37	38

Note. N = 120.

Intercorrelations Among Dimensions. For each method, the intercorrelations between the dimensions are presented in the monomethod triangles of the MTMM matrix (Table 7).

The correlation between C^2 and communications is moderate in size and significant for the field test ($p < .01$); it is near zero and insignificant on the SIMNET test. These results are similar to the first subset of items from the field and SIMNET tests and indicate a different factor structure between the dimensions as measured by the SIMNET measure compared to the field measure.

Multitrait-Multimethod Matrix. The multitrait-multimethod matrix of direct-analogue items from the field and SIMNET tests is presented in Table 7.

Table 7.

Multitrait-multimethod correlation matrix for the direct-analogue items of the field and SIMNET tests.

Method/ trait	FIELD		SIMNET	
	C^2	CO	C^2	CO
FIELD				
C^2				
CO	.32**			
SIMNET				
C^2	**.17***	.01		
CO	.04	**.31****	.03	

Note. The correlations in the heteromethod block with boldface type are convergent validity coefficients. The solid triangles are heterotrait-monomethod correlations. The dashed triangles are heterotrait-heteromethod correlations. C^2 = Command and Control, CO = Communications. *$p < .05$, **$p < .01$. Underlined correlations are negative.

The convergent validity coefficients for both dimensions are significant but low. Discriminant validity is supported by the heterotrait-heteromethod correlations which are lower than the convergent validity coefficients. However, the correlation between C^2 and communications is higher on the field test than the convergent validity coefficient for C^2 which does not support discriminant validity.

The convergent validity coefficients were corrected for unreliability in both measures. The corrected correlations are .23 and .46 for C^2 and communications, respectively. This indicates little improvement in convergence between the two methods for these dimensions even with perfectly reliable measurement.

Analysis of Variance. The ANOVA results for the direct-analogue items of the field and SIMNET tests are presented in Table 8. The results of the significance tests on the main effect and interactions indicate that each source of variance is significant. The significant Subject variance indicates that there is differentiation among test crews attributable to the test (convergent validity). There is also differential ordering of test crews on the different performance dimensions as indicated by the significant Subject X Dimension interaction (discriminant validity); its variance component is larger than the Subject effect. The Subject X Method interaction is also significant but is the weakest of the three effects. The analysis of variance procedure provides some evidence of convergent validity, good evidence of discriminant validity, and some method bias. However, the large error variance component indicates that more variance is attributable to unknown sources than to desirable sources.

Table 8.

Analysis of variance for direct-analogue items.

Source	df	MS	F	Variance Component
Subject (S) (test crews)	119	1.43	2.21*	.1947
S X dimension	119	1.07	1.65*	.2642
S X method	119	.88	1.54**	.1690
Error	119	.65		.5887

Note. * $p < .001$, ** $p < .025$.

The ANOVA results for the direct-analogue items of the field and SIMNET tests appear more supportive for the construct validity of SIMNET than the results using the larger subset of items. However, it is not appropriate to make direct comparisons of the ANOVA result between the two subsets of items. The ANOVA for the larger subset of items included the correlations for two more performance dimensions than the ANOVA for the direct-analogue items. Evidence for convergent and discriminant validity was lacking for these two dimensions, position location and combat driving, as shown through the MTMM matrix. This would also be reflected through the ANOVA procedure. Since the ANOVA on the direct-analogue items did not include these dimensions it would seem logical for the convergent and discriminant validity results to be more supportive.

Comparison of Corrected Convergent Validity Coefficients

The corrected convergent validity coefficients for the two subsets of items for the dimensions of C^2 and communications were compared to determine if the direct-analogue items produced higher correlations than the larger subset of items. The opposite was true as illustrated in Table 9.

Table 9.

Comparison of corrected convergent validity coefficients.

DIMENSION	FIRST SUBSET	DIRECT-ANALOGUE
Command and Control	.26	.23
Communications	.56	.46

Note. N = 120.

Task Similarity Questionnaire

The means and standard deviations for the TSQ items and composites are presented in Table 10. Opinions on the similarity of performance of items is on the average "mostly the same" indicating that the soldiers felt that performing those tasks on SIMNET was "mostly the same" as performing those tasks on the M1 tank in the field. The dimension composites are also included on Table 10: dimension 1 = C^2, dimension 2 = communications, dimension 3 = position location, and dimension 4 = combat driving.

Table 10.

Means and standard deviations for the Task Similarity Questionnaire items and composites.

	Dimension	Mean	SD
Giving combat reports	2	4.33	.95
Following radio procedures	2	4.44	.88
Commanding the crew	1	4.29	.90
Directing engagements	1	4.05	1.03
Determining position location	3	3.30	1.25
Map reading	3	3.36	1.28
Combat driving	4	3.36	1.26
Security		3.55	1.18
Call and adjust indirect fire		4.09	1.10
Troop leading procedures		4.23	.92
Grand Mean		3.89	1.08
C^2 Composite		4.17	.84
CO Composite		4.38	.81
PL Composite		3.33	1.13
CD Item		3.36	1.26

Note. N = 240. C^2 = Command and Control, CO = Communications, PL = Position Location, CD = Combat Driving. SD = standard deviation.

Comparison of TSQ Composites and Subset One Convergent Validity Coefficients

The rank order of the convergent validity coefficients from the first subset of items from the field and SIMNET tests and the mean similarity score on the four dimensions is presented in Table 11.

Table 11.

Rank order of field and SIMNET test convergent validity coefficients and TSQ composites mean similarity score.

	Convergent Validity Coefficients	Rank	TSQ	Rank
Command and Control	.20	2	4.17	2
Communications	.43	1	4.38	1
Position Location	.09[a]	3	3.33[b]	4
Combat Driving	-.03[a]	4	3.36[b]	3

Note. [a] These coefficients are not significantly different from zero and therefore essentially = zero, therefore their rankings could easily be reversed. [b] These ratings are not statistically different ($t = .40$, $p > .05$).

Rankings are consistent with communications receiving the highest ranking and C^2 the second highest. Position location and combat driving convergent validity coefficients and mean similarity score are very similar and therefore they virtually tie for the third and fourth place ranking.

Chapter V

Discussion

The purpose of this research was to gather evidence to evaluate the construct validity of four SIMNET measures of Armor combat performance by examining their relationship to field measures. The data suggest that SIMNET's construct validity cannot be defended on this basis.

The average level of performance on the dimensions was similar across the two methods, field and SIMNET tests. Also, SIMNET appears capable of reliably measuring performance on the critical domains when a sufficient number of items are included in the test. However, the crucial question is the degree to which scores on the two methods correlate with one another. Clearly, crew performance on the SIMNET measures for position location and for combat driving skills do not correspond to their performance on the field measures as evidenced by the non-significant convergent validity coefficients. In one sense, this is not too disturbing considering SIMNET's developers intended it to be used to train and test C^2 and communications skills and they recognize its limitations regarding combat critical skills (Pope, 1987). The convergent validity coefficients were significant for the C^2 and communications dimensions, but low. Greater convergent validity must be demonstrated before SIMNET can be considered a good surrogate testing method for the field exercise.

Although there is evidence of discriminant validity for the C^2 and communications dimensions, the dimensions intercorrelate differently on the two methods. The interrelationships among the dimensions on the field test are more explainable. The correlation between C^2 and communications is the highest. These tasks are all performed by the TC and are sometimes referred to collectively as command, control, and communications (C^3) skills because of their similarity. However, the inter-correlations among the SIMNET dimensions is perplexing. The position location and combat driving dimensions were most highly intercorrelated. The TC performs the position location tasks while the driver performs the combat driving tasks. There does not appear to be a logical explanation for the high relationship between these two dimensions. Replication of this finding appears necessary before we invest further effort to understand the causes of these unexpected results.

The ANOVA procedure demonstrated the large amount of method bias as its variance contribution was greater than the desirable effect attributable to discriminant validity and was close to the desirable effect attributable to convergent validity. The large amount of error variance indicated more variance was due to unknown sources than to the desirable effects.

Explanations for Low Convergence

There are several possible explanations for low convergence between the two methods. It is possible that using subjects who were previously naive to SIMNET could have restricted the validity coefficients. Although subjects were trained as well as the available resources allowed, differential rates of learning on the device could have influenced the crew's test scores. For example, there could be two crews who perform equally well on the field test, but because the TC of one of the crews learns slowly on new devices his crew does poorly on the SIMNET test while the other crew performs about the same as they did on the field test. However, if the slow learning TC had more exposure to SIMNET, his crew may have done as well on the SIMNET test as the other crew. Ideally, subjects should be experienced on both measures to eliminate the effects of differential rates of learning so that their test scores would be more stable and produce higher validity coefficients.

Another possible reason for low convergence between the two methods is that items and scenarios were not the same across the tests. To investigate the assumption that comparing performance on the same set of items would produce higher convergent validity coefficients, analyses were repeated using only direct-analogue items. The direct-analogue subset of items appeared to be measured reliably with fewer numbers of items in the dimensions. But most importantly, the corrected validity coefficients were not higher than those obtained with the "less analogous," larger subset of items. Also, the ANOVA results did not provide compelling evidence of convergent and discriminant validity for the direct-analogue items. Therefore, it appears that using a subset of items that more broadly taps the constructs, as was done in this research, did not restrict the correlations between performance on the measures.

The scenarios surrounding the performance of the test items, including the direct-analogue items, were different for the two tests and this remains a possible cause for the low convergent validity coefficients. For example, SIMNET was a platoon-level exercise and therefore tapped skills of the TC relating to platoon functions, such as forming tank formations and platoon communications requirements. However, the field test being a single tank exercise did not require platoon-level skills. In this respect more experienced crews, such as those with a TC who had been a platoon sergeant, were likely to do better on the SIMNET test when platoon-level skills were tapped. However, these more experienced crews would not have an advantage on the field test since these skills were not required. The emphasis on platoon versus individual tank skills on the

SIMNET and field tests, respectively, would tend to lower correlations of performance on the two tests.

Another possible reason for low convergence may be that the test scores derived for the field and SIMNET tests are unstable over time. Internal consistency reliability was determined through this research as is appropriate for a construct validity assessment. The convergent validity coefficients for C^2 and communications were corrected for attenuation according to the internal consistency of the dimensions. The corrected convergent validity coefficients were not very much higher than the uncorrected coefficients because the internal consistency for these two dimensions on both tests was good to start with. However, when considering the use of test scores for soldier evaluation purposes, the stability of those scores over time is also of critical interest. The stability of test scores is usually measured through correlating subjects' scores on the test at one point in time with their scores on the test at another point in time. It was not possible to assess test-retest reliability in the present research. If the SIMNET and field tests are not very reliable over time than the convergent validity coefficients would be attenuated. Future research should examine test-retest reliability and use this reliability estimate to correct the obtained validity coefficients.

Another problem involved one TC who indicated his dislike for simulation devices because he perceived an effort by simulation advocates to eliminate field exercises. His crew scored close to the highest on the field test, yet he purposely did poorly on the SIMNET test, on his own admission, and as a result his crew scored close to the bottom on the SIMNET test. Although this was the only case where an indication was given that a subject purposely performed poorly on SIMNET, his attitude could have been shared by others and manifested in either a conscious or unconscious effort to "make SIMNET look bad" by performing poorly.

Problems with Criterion Relevance

It was discussed previously in this thesis that the field exercise is accepted by the Armor training and testing community as the most realistic and relevant substitute for the ultimate criterion of combat performance. That is why a field test was used to assess the construct validity of SIMNET. However, it is possible that the field test itself is not a good measure of combat performance. Although the relevancy of testing on the actual equipment in a field setting seems apparent, there are many ways in which SIMNET may be a better measure of combat performance than a field test.

SIMNET provides a great variety of testing conditions that can not possibly be tested in the field because of monetary and safety restraints. For example, SIMNET allows soldiers to fight other soldiers so that there is a very real threat from enemy fire. Once a tank is disabled from enemy fire, it is out of commission for the rest of the exercise. Tanks in SIMNET must also deal appropriately with terrain features or risk disablement. This could result from the tank driving off a cliff or getting stuck in an unfordable stream. These situations could happen easily on an actual battlefield but these conditions would not generally be present on a field test course for obvious safety and tank maintenance reasons. Also, engagements from enemy helicopters and Anti-tank guided missiles for example, can be simulated through SIMNET, but not in the field.

Another realistic feature of SIMNET is its ability to train or test up to the brigade-level. Field exercises at this level are rarely performed for training purposes because of the high cost involved and it would be practically impossible to obtain the funding for research purposes. Yet, tanks in battle would generally be found acting as part of a platoon rather than individually. This increases the complexity of the C^3 skills which can be evaluated in SIMNET compared to a field measure.

SIMNET also allows tanks to traverse cross country as opposed to following a prescribed course. Field tests generally do not allow cross country movement because of the controls necessary for standardization and data collection.

In an actual battle, tanks would rarely be following a road.

The field test used in this research was better than most in terms of realism and variety of conditions, but it was still limited by the above considerations. For example, the tanks acted independently, other than the radio communication with higher headquarters, and followed a prescribed course along a road. Also, engagements were restricted to stationary tanks and infantry.

The SIMNET test was designed to take advantage of SIMNET's capabilities. It incorporated engagements with helicopters and anti-tank guided missiles and platoon formations into a cross country, platoon-level exercise with a wide variety of terrain features. Despite the fact that the soldiers were not in an actual tank, in some respects the SIMNET test was much closer to simulating an actual battle.

The soldiers participating in this research supported the face validity of SIMNET. Their questionnaire responses indicated that performing tasks on SIMNET was "mostly the same" as performing tasks on the M1 tank. Soldiers did discriminate among tasks in terms of the degree of similarity between the behavior required for task performance by SIMNET and the M1 tank. Likewise, the convergent validity coefficients for the four dimensions were also of varying strengths. The rank order of soldiers opinions on the TSQ composites and the rank order of convergent validity coefficients were essentially the same indicating correspondence between soldiers opinions and performance.

It is possible that SIMNET is a better measure of combat performance than the field test because of SIMNET's ability to incorporate realistic battle conditions without safety or monetary concerns. If so, then this could be one reason for the lack of convergence between the field test and the SIMNET test. This would indicate that a field test is not the appropriate criterion to compare SIMNET performance. It would seem that actual combat performance would be needed against which to assess SIMNET's construct validity. However, without actual combat performance measures, this question can not be addressed and the Armor community, in general, remains supportive of field testing as a relevant and realistic measure of combat performance.

Conclusion

In conclusion, this research does not support the construct validity of SIMNET as a performance testing device of combat critical skills when those skills are operationally defined by field test performance. It is important for more research to be done on this topic since SIMNET is currently being used in this capacity and its proponents contend it is useful for this purpose. Future research should use SIMNET experienced soldiers. This will become more of a possibility as more soldiers are being exposed to SIMNET for training and developmental purposes. In the future an effort should be made to construct SIMNET test items and scenarios that mirror their field test manipulations as closely as possible. If strong correlations are still not found between SIMNET and field performance with these conditions held constant, then a search for a better criterion against which to assess SIMNET's construct validity may be necessary.

Of course, it remains a possibility that SIMNET is not a valid performance evaluation device of combat critical skills and future research would fail to support SIMNET for this purpose. Yet, considering SIMNET's proven utility for training and developmental purposes and face validity as a performance evaluation device, further research incorporating the above recommendations appears warranted before SIMNET's validity can be supported or disputed.

Bibliography

Black, B. A., and Mitchell, K. J. (1986). Predicting performance of M1 gunners. (ARI Technical Report 707). Fort Knox, KY: U.S. Army Research Institute.

Boldovici, J. A. (1981). Some problems in evaluating training devices and simulators. (HumRRO-PP-2-81). Alexandria, VA: Human Resources Research Organization.

Bolt, Beranek, and Newman Software Products Corporation. (1987). RS/1 user's guide. Cambridge, MA: Author.

Brown, F. J. (1983). The use of simulation of armor unit tactical training. Paper presented at NATO Armor School Commander's Conference, Samur, France.

Brown, R. E., Pishel, R. G., and Southard, L. D. (1988). Simulation Networking - Preliminary training developments study. (TRAC-WSMR-TEA-8-88). White Sands, NM: U.S. Army TRADOC Analysis Command.

Campbell, D. T., and Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. Psychological Bulletin, 56, 81-105.

Cascio, W. F. (1982). Applied psychology in personnel management (2nd ed.). Reston, VA: Reston Publishing Company, Inc.

Chung, J. W., Dickens, A. R., O'Toole, B. P., and Chiang, C. J. (1988). SIMNET M1 Abrams main battle tank simulation: Software description and documentation (Revision 1). Cambridge, MA: Bolt, Beranek, and Newman Systems and Technologies Corporation.

DuBois, R. S. (1987). The M1 Unit-Conduct of Fire Trainer (U-COFT) as a tank gunnery testing device: A psychometric evaluation. Unpublished master's thesis, Western Kentucky University, Bowling Green, KY.

Dubois, R. S., and Smith, P. G. (1989). A simulation-based evaluation of a position navigation system for Armor: Soldier performance, training, and functional requirements. (ARI Technical Report 834). Fort Knox, KY: U.S. Army Research Institute.

Garvey, R. E. and Radgowski, T. (1988). SIMNET-D standard operating procedure. (Report No. 6929). Cambridge, MA: Bolt, Beranek, and Newman Systems and Technologies Corporation.

Ghiselli, E. E., Campbell, J. P., and Zedeck, S. (1981). Measurement theory for the behavioral sciences. San Francisco, CA: W. H. Freeman and Company.

Goldstein, I. L. (1986). Training in organizations (2nd ed.). Monteray, CA: Brooks/Cole Publishing Company.

Gound, D. and Schwab, J. (1988). Concept evaluation program of Simulation Networking (SIMNET). (TRADOC TRMS No. 86-CEP-0345). Fort Knox, KY: U.S. Armor and Engineer Board.

Graham, S. E. (1986). The Unit-Conduct of Fire Trainer (U-COFT) as a medium for assessing gunner proficiency. (ARI Research Report 1422). Fort Knox, KY: U.S. Army Research Institute Report.

Graham, S. E., Leet, W. T., Elliot, G. S., Hamill, J. P., and Smith, S. E. (1989). Soldier performance research project: Armor field and SIMNET tests. (ARI Research Report 1454). Fort Knox, KY: U.S. Army Research Institute.

Hoffman, R. G. and Morrison, J. E. (1988). Requirements for a device-based training and testing program for M1 gunnery: Volume 1. Rational and summary of results. (ARI Technical Report 783). Fort Knox, KY: U.S. Army Research Institute.

Kavanagh, M. J., MacKinney, A. C., and Wolins, L. (1971). Issues in managerial performance: Multitrait-multimethod analysis of ratings. Psychological Bulletin, 75, 34-49.

Kraemer, R. E. and Bessemer, D. W. (1987). U.S. tank platoon training for the 1987 Canadian Army Trophy (CAT) competition using a Simulation Networking (SIMNET) system. (ARI Research Report 1457). Fort Knox, KY: U.S. Army Research Institute.

Lawler, E. E. (1967). The multitrait-multirater approach to measuring managerial job performance. Journal of Applied Psychology, 51, 369-381.

Pavitt, C. and Tomich, L. (1982). Characteristics and description book: M1 tank. Fort Knox, KY: General Dynamics.

Perceptronics, Incorporated. (1987). SIMNET M1 crew manual. Woodland Hills, CA: Author.

Pope, A. R. (1987). The SIMNET network and protocols. (Report No. 6369). Cambridge, MA: Bolt, Beranek, and Newman Laboratories Incorporated.

Schwab, J. R. (1987). Innovative test of Simulation Networking - Developmental (SIMNET-D). (TRADOC TRMS No. 87-0000778). Fort Knox, KY: U.S. Armor and Engineer Board.

Smith, N. D. (1978). State of the art: OPFOR and ARTEP implementation in the U.S. Army. (ARI Research Problem Review 78-25). Fort Hood, TX: U.S. Army Research Institute.

Thomson, H. A. (1970). Comparison of predictor and criterion judgments of managerial performance using the multitrait-multimethod approach. Journal of Applied Psychology, 54, 496-502.

U.S. Department of the Army (1984). M1 tank combat tables (FM 17-12-1). Washington, D.C.: Author.

U.S. Department of the Army (1988). Mission training plan for the platoon (ARTEP 17-237-10-MTP). Washington, D.C.: Author.

Appendix A

FIELD TEST EVENTS

Station 1
Brigade Support Area

The brigade support area (BSA) was a tactical station which replicated, as closely as possible, a portion of a BSA in a combat situation. The crew members met each other for the first time and were told to prepare an M1 tank for combat. The TC was given the mission of taking the tank forward to a battalion currently in contact. The tank required ammunition upload, refueling, preventative maintenance checks and services (PMCS), and prepare-to-fire checks. There were four induced faults in the vehicle that the TC had to find and correct. In addition, the TC had to assist the gunner in preparing his station. The TC was required to conduct communications checks, enter a radio net, post an overlay, and review his orders with the crew. The operations order required that the time spent in the BSA was approximately two hours.

Station 2
Surprise Engagement with Disabled T72 and T72 in Overwatch

At a designated point in the road, the surrogate loader identified two tanks to be engaged at about 1200 meters. The TC was required to lay the main gun on the overwatch tank (most dangerous target) and give proper fire commands for the engagement. When the first T72 was engaged, it gave a visual signature that it has been hit (fire extinguisher smoke). The crew was then to engage the second T72. The second T72 gave an indication of having been hit after the first round was fired and its crew evacuated the vehicle and ran into the woods. The driver should have then turned the frontal armor toward the targets, terrain permitting. The TC was required to engage both crews with his machine gun. The TC should have reported the action to his higher headquarters giving a correct location and directing his crew to assume a battlecarry posture with SABOT loaded.

Station 3
ATGM Ambush in Minefield

The TC was required to correctly locate the minefield from an overlay he was given at the BSA. The TC should have directed the driver to a cleared and marked lane through the minefield and control the driver's progress through it. As the tank approached a point at about one-third through the minefield, it was engaged by an anti-tank guided missile (ATGM) from a vehicle partially concealed 1500-2000 meters to the direct front. The gunner was to acquire the ATGM blast and alert the TC who should have immediately issued a fire command against the OPFOR vehicle. The TC should then have directed the driver to rapidly move forward out of danger firing at the OPFOR vehicle with the main gun and/or TC's machine gun. The gunner continued to engage until the TC determined the target was destroyed. Another possible solution to the situation was for the TC to direct the driver to move rapidly backward, activating vehicle smoke. In 15 to 20 seconds when the smoke had sufficiently cleared, the TC would lay the main gun on the target and continue to engage until destroyed. The TC would then direct the proper battlecarry posture and submit a correct report.

Station 4
Meeting Engagement with Enemy Stragglers: Loader Killed

At this station, the TC acquired three enemy soldiers at approximately 40 meters about the same time the enemy soldiers opened fire on the tank with automatic rifle fire. The loader was killed. The loader had a bag of fake blood which he squirted over the inside of the tank and the TC in order to make his death convincing and stressful to the TC. The TC engaged the enemy with the coax machine gun, or directed the gunner to engage the enemy with his machine gun. The TC checked the loader and determined him to be dead. The TC should have then submitted a correct report to his higher headquarters and requested instructions. He was told to leave the loader by the side of the road and he would be picked up later. The TC, gunner, and driver were to evacuate the loader to the side of the trail, prepare the tank for operation in a three-man crew configuration, and proceed.

Station 5
Military Police Traffic Check Point

As the tank approached the traffic check point (TCP), the TC should have recognized the TCP as friendly military police (MP). The TC stopped the tank, and the MP checked the TC's navigation. The TC then proceeded, according to the MP's directions, toward the correct location.

Station 6
Meeting Engagement with T72 and BMP at Short Range

A T72 leading an enemy infantry vehicle, or BMP, appeared heading the opposite way along the route of the tank at short range (under 500 meters). As soon as the TC acquired the T72 he should have layed on the main gun, announced "On the Way", and fired. After the first round was fired, there was no indication that the target had been hit. The TC was required to re-engage the T72. The TC was then to engage the BMP as it unmasks from behind the T72. The BMP was destroyed on the first round. The TC should have then directed the correct battlecarry posture and submitted a correct report of the action.

Station 7
Automatic Weapons Ambush: TC and Gunner Killed

A close range (100 meters) automatic weapons ambush occurred in which the TC was immediately killed. The loader (formerly the gunner) was able to communicate to the driver that the TC was killed and he was hit and losing consciousness. At this point, the driver, under his own initiative, was to move the tank out of the kill zone, determine crew status, submit a report giving vehicle location, and report casualties. The driver was then directed to proceed. He was stopped at the end of the lane (a short time later) by controllers. He was then required to correctly identify his unit, mission, and determine his location.

FIELD TEST TASK LIST

	TASK	PERFORMANCE DIMENSION
1.	Minefield plotted on map	PL
2.	Plot matches decoded coordinates	PL
3.	TC lays on most dangerous target	C^2
*4.	Proper fire command elements "Gunner"	C^2
*5.	"Sabot" (or Battlesight)	C^2
*6.	"Two tanks"	C^2
*7.	"Right tank"	C^2
8.	Waits for "Up" & "Identified"	C^2
*9.	"Fire and adjust"	C^2
10.	Drives at constant speed or seeks hull-down	CD
*11.	Submits report without being cued	COMMO
*12.	Elements of report -Correct call sign	COMMO
*13.	Type of report: "Spotrep"	COMMO
*14.	What happened: "Destroyed two T72s"	COMMO
15.	Grid: (+/- 200 meters)	PL
16.	Correct "Time"	COMMO
*17.	What you are doing: "Continuing Mission"	COMMO
18.	TC directs driver to use cleared lane	C^2
19.	TC directs driver through minefield or dismounts loader	C^2
20.	Vehicle visibly stays in cleared lane	CD
21.	TC directs driver to speed up or backup and engage smoke	C^2

* Indicates Subset Two Items

22.	Driver protects tank after ATGM is launched	CD
*23.	Proper fire command elements "Gunner"	C^2
*24.	"Sabot" (or Battlesight)	C^2
*25.	"PC" (or BMP)	C^2
*26.	"Fire"	C^2
*27.	"Fire Heat"	C^2
*28.	Submits report without being cued	COMMO
*29.	Elements of report -Correct call sign	COMMO
*30.	Type of report: "Spotrep"	COMMO
*31.	What happened: "Destroyed 1 BMP"	COMMO
32.	Grid: (+/- 200 meters)	PL
*33.	Correct "Time"	COMMO
*34.	What you are doing: "Continuing Mission"	COMMO
*35.	Proper fire command elements "Coax"	C^2
*36.	"Troops"	C^2
*37.	"Fire and Adjust"	C^2
*38.	"Caliber .50"	C^2
39.	Driver positions tank appropriately	CD
40.	TC moves gunner to loader's position	C^2
41.	TC prepares weapon station	C^2
*42.	Submits report without being cued	COMMO
43.	Elements of report -Correct call sign	COMMO
*44.	Type of report: "Spotrep"	COMMO
*45.	What happened: "Destroyed"	COMMO
46.	Grid: (+/- 200 meters)	PL
*47.	Correct "Time"	COMMO

*48.	What you are doing: "Continuing Mission"	COMMO
*49.	Submits casualty report without being cued	COMMO
*50.	Elements of report -Personnel battle loss report or "Red 2"	COMMO
51.	Identifies correct battle roster number	COMMO
*52.	Correct "Date/Time"	COMMO
53.	"4A"	COMMO
54.	"Left body on tank"	COMMO
55.	Identifies correct grid (+/- 200 meters)	PL
56.	Identifies route on map correctly	PL
57.	Takes correct turns in route to Station 6	PL
*58.	Proper fire command elements "Battlesight"	C^2
*59.	"Tank"	C^2
*60.	Waits for "Up"	C^2
*61.	"On the way"	C^2
62.	Driver protects tank	CD
63.	TC announces "On The Way"	C^2
64.	TC or Driver announces "Target"	C^2
65.	TC engages BMP "On the Way"	C^2
66.	TC or Driver announces "Target"	C^2
67.	TC engages troops with CAL .50	C^2
68.	Proper fire command "Caliber .50"	C^2
*69.	Submits report without being cued	COMMO
*70.	Elements of report -Correct call sign	COMMO
*71.	Type of report: "Spotrep"	COMMO
*72.	What happened: "Destroyed T72 and PC"	COMMO
73.	Grid: (+/- 200 meters)	PL

*74.	Correct "Time"	COMMO
*75.	What you are doing: "Continuing Mission"	COMMO
76.	Driver protects tank	CD
77.	Driver submits report without being cued	COMMO
78.	Elements of report -Correct call sign	COMMO
79.	Type of report: "Spotrep"	COMMO
80.	What happened: "Four to Six/Infantry Ambush"	COMMO
81.	What you are doing: "Continuing Mission"	COMMO
82.	Submits casualty report without being cued	COMMO
83.	Elements of report -Identifies TC as casualty	COMMO
84.	Identifies gunner as casualty	COMMO
85.	Driver locates his position	PL

Appendix B

SIMNET TEST EVENTS

Event I
Crew Joins Platoon as Wingman

The TC received an order and entered the platoon radio net. The crew then operated as part of a tank platoon during a tactical road march. The TC was required to properly supervise the positions of the tank during movement and short halts. At the direction of the platoon leader, the platoon assumed several formations such as the coil, herringbone, and vee. The tank was to move tactically as the wingman for the platoon sergeant (PSG). When told, the crew should have properly executed an action drill by orienting the main gun in the proper direction and maintaining movement, orientation, and position. Shortly thereafter, the tank was to perform an air attack drill. The TC was then to issue a proper fire command. The TC was asked by the PSG to determine the platoon's location.

Event II
Platoon Encounters Bridge

The platoon formation encounted a bridge. The driver was to maintain the proper position with respect to the PSGs tank and the proper overwatch. The TC must have determined the location of the bridge and send a spot report (SPOTREP) stating that they were crossing the bridge and give the bridge's correct location.

Event III
Three T72s are Observed

The platoon conducted a meeting engagement with an enemy tank platoon. The PSG acquired the targets, directed a contact drill, and asked the TC to issue a contact report. The crew then began an action drill. The driver was required to maintain proper position. The tank should have then used proper engagement priorities. When all enemy tanks had been destroyed, the TC should have sent a SPOTREP reporting their activity and location. Then the friendly platoon resumed movement during which execution of section formations and drills were evaluated.

Event IV
Enemy ATGM Attacks Formation

The platoon was attacked by helicopters. During the attack, the PSG's tank was destroyed. The tank should have engaged the helicopter, issued a contact report, executed a contact drill, and conducted an air attack drill. The platoon resumed movement with the tank now assuming the PSG position in the platoon. The TC was required to send a situation report stating the action encountered, casualties, location, and their new position in the platoon formation. The platoon resumed in a vee formation and the crew was evaluated on the execution of that formation and drills.

Event V
Reaction to ATGM Ambush

The crew reacted to an ATGM ambush, The TC was required to issue the contact report and fire command. The tank should have taken evasive action (TC and driver responsibility) and engaged the enemy until the enemy was destroyed. The TC was required to submit a proper SPOTREP.

Event VI
React to Indirect Fire

The crew reacted to indirect fire by speeding through the area. The TC was then required to give a SPOTREP describing what happened and the proper location of the activity.

Event VII
Engagement From Hasty Battle Position

The crew was required to assume a hasty fighting position and engage a reinforced motorized rifle company (MRC) as part of the platoon. The platoon leader issued a platoon fire command. As part of the platoon, the crew unmasked from a hill top and engaged the MRC. The MRC was in platoon columns approximately 2,500 meters in front of the fighting position. As the MRC was taken under fire, it returned fire and moved into a company line to assault the fighting position. All the enemy tanks were destroyed. The other friendly tank had a mobility failure (shears a sprocket) in a partially exposed position. The three surviving BMPs from the MRC took effective cover approximately 1,500 meters to the front of the fighting position. The test crew was aboard the only undamaged tank remaining in the platoon. The TC was required to enter the company radio net and report. The TC requested instructions.

Event VIII
Request and Adjust Indirect Fire

The Company Commander sent coordinates of other platoons. The tank could not take the targets under effective direct fire. The TC was required to call for and adjust indirect fire on the target.

SIMNET TEST TASK LIST

	TASK	PERFORMANCE DIMENSION
1.	Maintains visual contact with PSG's tank	CD
2.	Maintains position 100-150 meters from PSG's tank	CD
3.	Takes up position on opposite side of column from PSG's tank	CD
4.	Maintains correct gun tube orientation	PL
5.	Driver orients vehicle at 3 o'clock position	CD
6.	Driver maintains gun tube orientation	PL
7.	Driver takes proper position	CD
8.	Driver pulls tank off route and stops	CD
9.	TC ensures gun covers the column's rear and tank is within sight of the other tanks	C^2
10.	Wingman takes proper position	CD
11.	Wingman maintains overwatch	CD
12.	Driver turns vehicle 90 degrees to left	CD
13.	Maintains visual contact with PSG	CD
14.	Driver takes proper position	CD
15.	Gun tube orientation	PL
16.	Makes sudden turns	CD
17.	Driver changes speed	CD
18.	Maintains proper gun tube orientation	PL
19.	Driver orients vehicle at 3 o'clock position	CD
20.	Gun tube orientation	PL
21.	Grid coordinates (+/- 200 meters)	PL

---

* Indicates Subset Two Items

22.	Driver maintains proper position	CD
23.	Proper overwatch	CD
24.	Sends Spot Report without cue	COMMO
25.	Elements of report -Grid coordinates (+/- 200 meters)	PL
26.	Activity "Crossing Bridge"	COMMO
*27.	"Continuing mission"	COMMO
28.	Contact Report elements "Contact"	COMMO
29.	"Three tanks"	COMMO
30.	"WEST"	PL
31.	Driver turns own tank toward enemy tank	CD
32.	Driver maintains proper position	CD
*33.	Proper fire command elements "Gunner"	C^2
*34.	"Sabot"	C^2
*35.	"Three tanks"	C^2
*36.	"Left tank first"	C^2
*37.	"Fire"	C^2
38.	Wingman bounds, maintains proper position	CD
39.	Engages until all tanks are destroyed	C^2
*40.	Sends report to platoon leader w/o cue	COMMO
*41.	Elements of report -Identifies "SPOTREP"	COMMO
*42.	Correct Call sign ("red 3")	COMMO
*43.	"Destroyed three T72s"	COMMO
44.	Number of rounds fired	COMMO
45.	Driver maintains proper position	CD
46.	Gun tube orientation	PL
47.	Contact Report elements "Contact"	COMMO

48.	"North"	PL
49.	"BMP"	COMMO
50.	Driver turns tank 45 degrees from attacking aircraft	CD
*51.	Proper fire command elements "Gunner"	C^2
*52.	"Sabot"	C^2
*53.	"PC"	C^2
*54.	"Fire"	C^2
*55.	"Fire Heat"	C^2
56.	Elements of report -Correct call signs	COMMO
*57.	Type of report: "Sitrep"	COMMO
*58.	Correct DTG	COMMO
*59.	"Destroyed enemy BMP"	COMMO
60.	Grid: (+/- 200 meters)	PL
61.	Line 4: "Correct"	COMMO
62.	Line 5: "None"	COMMO
63.	Line 6: "Red"	COMMO
64.	Correct ammo status	COMMO
65.	Correct fuel status	COMMO
*66.	"Continuing mission"	COMMO
67.	TC assumes proper position	C^2
68.	TC maintains visual contact	C^2
69.	Proper gun tube orientation	PL
70.	Contact report elements "Contact"	COMMO
71.	"Northeast"	PL
72.	"Missile"	COMMO
73.	Driver takes evasive action	CD

74.	Proper fire command elements "Gunner"	C^2
*75.	"Sabot"	C^2
*76.	"PC"	C^2
*77.	"Fire"	C^2
*78.	"Fire Heat"	C^2
*79.	Submits report without cue	COMMO
80.	Elements of report -Correct call signs	COMMO
*81.	Type of report: "Spotrep"	COMMO
*82.	What happened: "Destroyed BMP"	COMMO
83.	Grid: (+/- 200 meters)	PL
*84.	Correct "Time"	COMMO
*85.	What you are doing: "Continuing Mission"	COMMO
*86.	Submits report without cue	COMMO
*87.	Elements of report -Type of report: "Spotrep"	COMMO
88.	What happened: "Observing Indirect Fire"	COMMO
89.	Grid: (+/- 200 meters)	PL
*90.	Correct "Time"	COMMO
91.	Proper fire command elements "Gunner"	C^2
*92.	"Sabot"	C^2
*93.	"Tanks"	C^2
*94.	"Rear tank"	C^2
*95.	"Fire"	C^2
96.	Fires at rear tanks first, works forward	C^2
*97.	Submits report without cue	COMMO
*98.	Elements of report -Correct call sign	COMMO
*99.	Type of report: "Spotrep"	COMMO

*100.	"Engaged (Correct #) Tanks and BMPs"	COMMO
101.	Grid: (+/- 200 meters)	PL
*102.	Correct "Time"	COMMO
*103.	What you are doing: "Continuing Mission"	COMMO
104.	Proper fire command "Gunner"	C^2
105.	"Sabot"	C^2
106.	"Tanks"	C^2
107.	"Left tank"	C^2
108.	"Fire"	C^2
109.	TC engages left tank first	C^2
110.	TC directs fire to move left to right	C^2
*111.	Submits report without cue	COMMO
*112.	Elements of report -Correct call sign	COMMO
113.	Type of report: "Spotrep"	COMMO
114.	"Engaged or Destroyed (Correct number)"	COMMO
115.	Grid: (+/- 200 meters)	PL
116.	Correct "Time"	COMMO
117.	What you are doing: "Continuing Mission"	COMMO
118.	Contacts company commander without cue	COMMO
119.	Elements of report -Type of report: "Sitrep"	COMMO
120.	DTG	COMMO
121.	What happened: "Engaged two enemy company sized-units"	COMMO
122.	Grid: (+/- 200 meters)	PL
123.	"Line 4d; one/Red 1 destroyed/Red 2 mobility kill/ I have assumed Red 1 duties"	COMMO
124.	"None"	COMMO

125.	"Black"	COMMO
126.	Ammunition "Black" Fuel "Black"	COMMO
127.	Requests instructions	COMMO
128.	TC contacts company FIST/CO	COMMO

B7, F3